Lecture 17

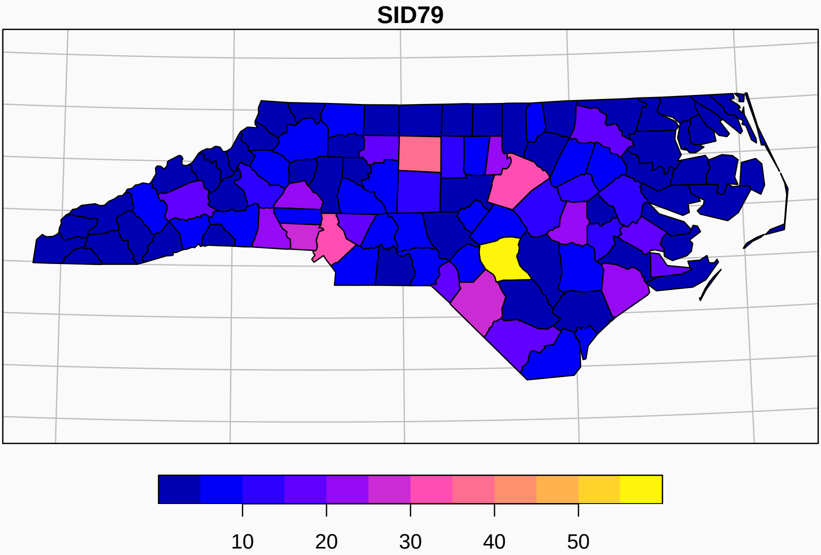
Models for areal data

Colin Rundel

11/02/2018

areal / lattice data

Example - NC SIDS



If we have observations at n spatial locations $(s_1, \ldots s_n)$

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \big(y(s_i) - \bar{y}\big)\big(y(s_j) - \bar{y}\big)}{\sum_{i=1}^n \big(y(s_i) - \bar{y}\big)^2}$$

where $\mathbf{w} = \{w\}_{ij}$ is a spatial weights matrix.

If we have observations at n spatial locations $(s_1, \ldots s_n)$

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y(s_i) - \bar{y})(y(s_j) - \bar{y})}{\sum_{i=1}^n \left(y(s_i) - \bar{y}\right)^2}$$

where $\mathbf{w} = \{w\}_{ij}$ is a spatial weights matrix.

Some properties of Moran's I when there is no spatial autocorrelation / dependence:

- $E(I) = -1/(n-1)$
- $Var(I) =$ Something ugly but closed form $- E(I)^2$
- $\lim\limits_{n\to\infty} \frac{I - E(I)}{\sqrt{Var(I)}} \sim \mathcal{N}(0,1)$

Adjacency Matrix

```r
1*st_touches(nc[1:12,], sparse=FALSE)
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
##  [1,]    0    1    0    0    0    0    0    0    0     0     0     0
##  [2,]    1    0    1    0    0    0    0    0    0     0     0     0
##  [3,]    0    1    0    0    0    0    0    0    0     1     0     0
##  [4,]    0    0    0    0    0    0    1    0    0     0     0     0
##  [5,]    0    0    0    0    0    1    0    0    1     0     0     0
##  [6,]    0    0    0    0    1    0    0    1    0     0     0     0
##  [7,]    0    0    0    1    0    0    0    1    0     0     0     0
##  [8,]    0    0    0    0    0    1    1    0    0     0     0     0
##  [9,]    0    0    0    0    1    0    0    0    0     0     0     0
## [10,]    0    0    1    0    0    0    0    0    0     0     0     1
## [11,]    0    0    0    0    0    0    0    0    0     0     0     1
## [12,]    0    0    0    0    0    0    0    0    0     1     1     0
```

Normalized Adjacency Matrix

```r
normalize_weights = function(w) {
  diag(w) = 0
  rs = rowSums(w)
  rs[rs == 0] = 1
  w/rs
}

normalize_weights( 1*st_touches(nc[1:12,], sparse=FALSE) )
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
##  [1,]  0.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.0
##  [2,]  0.5  0.0  0.5  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.0
##  [3,]  0.0  0.5  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.5   0.0   0.0
##  [4,]  0.0  0.0  0.0  0.0  0.0  0.0  1.0  0.0  0.0   0.0   0.0   0.0
##  [5,]  0.0  0.0  0.0  0.0  0.0  0.5  0.0  0.0  0.5   0.0   0.0   0.0
##  [6,]  0.0  0.0  0.0  0.0  0.5  0.0  0.0  0.5  0.0   0.0   0.0   0.0
##  [7,]  0.0  0.0  0.0  0.5  0.0  0.0  0.0  0.5  0.0   0.0   0.0   0.0
##  [8,]  0.0  0.0  0.0  0.0  0.0  0.5  0.5  0.0  0.0   0.0   0.0   0.0
##  [9,]  0.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0   0.0   0.0   0.0
## [10,]  0.0  0.0  0.5  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.5
## [11,]  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   1.0
## [12,]  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.5   0.5   0.0
```

NC SIDS & Moran's I

Lets start by using a normalized adjacency matrix for $\mathbf{w}$ (shared county borders).

```
morans_I = function(y, w) {
  w = normalize_weights(w)
  n = length(y)
  y_bar = mean(y)
  num = sum(w * (y-y_bar) %*% t(y-y_bar))
  denom = sum( (y-y_bar)^2 )
  (n/sum(w)) * (num/denom)
}

w = 1*st_touches(nc, sparse=FALSE)

morans_I(y = nc$SID74, w)
## [1] 0.1477405

ape::Moran.I(nc$SID74, weight = w) %>% str()
## List of 4
##  $ observed: num 0.148
##  $ expected: num -0.0101
##  $ sd      : num 0.0627
##  $ p.value : num 0.0118
```

Like Moran's I, if we have observations at n spatial locations $(s_1, \dots s_n)$

$$C = \frac{n-1}{2\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}\big(y(s_i) - y(s_j)\big)^2}{\sum_{i=1}^n \big(y(s_i) - \bar{y}\big)^2}$$

where $\mathbf{w}$ is a spatial weights matrix.

Like Moran's I, if we have observations at n spatial locations $(s_1, \dots s_n)$

$$C = \frac{n-1}{2\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}\big(y(s_i) - y(s_j)\big)^2}{\sum_{i=1}^n \big(y(s_i) - \bar{y}\big)^2}$$

where $\mathbf{w}$ is a spatial weights matrix.

Some properties of Geary's C:

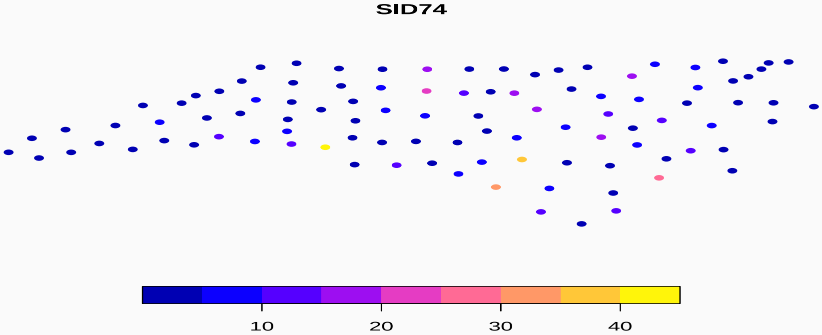
- $0 < C < 2$
  - If $C \approx 1$ then no spatial autocorrelation
  - If $C > 1$ then negative spatial autocorrelation
  - If $C < 1$ then positive spatial autocorrelation
- Geary's C is inversely related to Moran's I

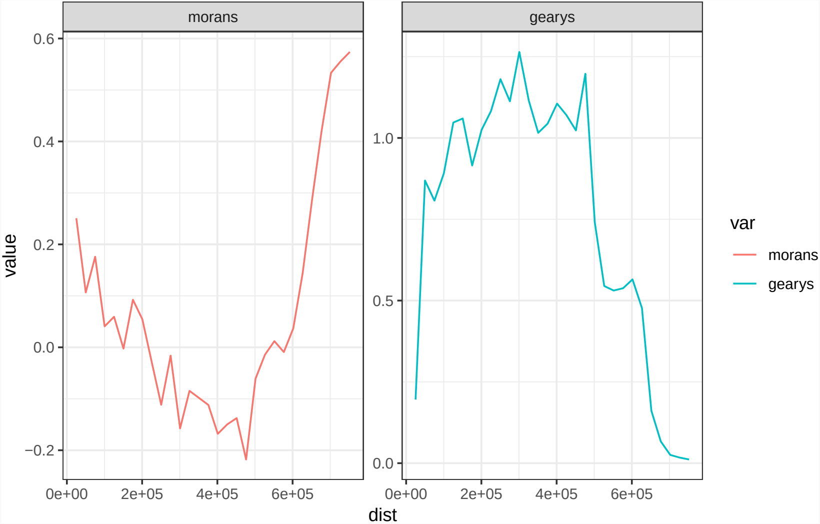
Again using an normalized adjacency matrix for $\mathbf{w}$ (shared county borders).

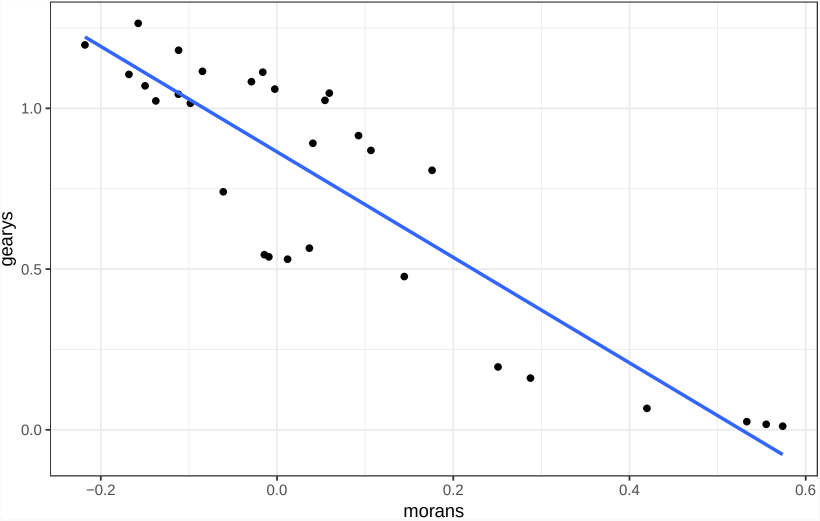
```
gearys_C = function(y, w) {
  w = normalize_weights(w)

  n = length(y)
  y_bar = mean(y)
  y_i = y %*% t(rep(1,n))
  y_j = t(y_i)
  num = sum(w * (y_i-y_j)^2)
  denom = sum( (y-y_bar)^2 )
  ((n-1)/(2*sum(w))) * (num/denom)
}

w = 1*st_touches(nc, sparse=FALSE)

gearys_C(y = nc$SID74, w = w)
## [1] 0.8438767
```

Spatial Correlogram

```
nc_pt = st_centroid(nc)
plot(nc_pt[,"SID74"], pch=16)
```




morans
gearys
value
dist
0.6
0.4
0.2
0.0
−0.2
1.0
0.5
0.0
0e+00
2e+05
4e+05
6e+05
var
morans
gearys



1.0
0.5
0.0
gearys
−0.2
0.0
0.2
0.4
0.6
morans

Autoregressive Models

Lets just focus on the simplest case, an $AR(1)$ process

$$y_t = \delta + \phi \, y_{t-1} + w_t$$

where $w_t \sim \mathcal{N}(0, \sigma^2)$ and $|\phi| < 1$, then

$$E(y_t) = \frac{\delta}{1-\phi}$$
$$Var(y_t) = \frac{\sigma^2}{1-\phi}$$
$$\rho(h) = \phi^h$$
$$\gamma(h) = \phi^h \frac{\sigma^2}{1-\phi}$$

AR Models - Time - Joint Distribution

Previously we saw that an $AR(1)$ model can be represented using a multivariate normal distribution

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \sim \mathcal{N}\left(\frac{\delta}{1-\phi} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},\ \frac{\sigma^2}{1-\phi} \begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{pmatrix} \right)$$

Previously we saw that an $AR(1)$ model can be represented using a multivariate normal distribution

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \sim \mathcal{N}\left(\frac{\delta}{1-\phi} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \; \frac{\sigma^2}{1-\phi} \begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{pmatrix} \right)$$

In writing down the likelihood we also saw that an $AR(1)$ is 1st order Markovian,

$$\begin{aligned} f(y_1, \ldots, y_n) &= f(y_1)\, f(y_2|y_1)\, f(y_3|y_2, y_1) \, \cdots \, f(y_n|y_{n-1}, y_{n-2}, \ldots, y_1) \\ &= f(y_1)\, f(y_2|y_1)\, f(y_3|y_2) \, \cdots \, f(y_n|y_{n-1}) \end{aligned}$$

Alternative Definitions for y_t

$$y_t = \delta + \phi\, y_{t-1} + w_t$$

vs.

$$y_t | y_{t-1} \sim \mathcal{N}(\delta + \phi\, y_{t-1},\ \sigma^2)$$

Alternative Definitions for y_t

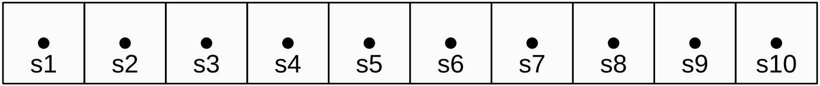
$$y_t = \delta + \phi\, y_{t-1} + w_t$$

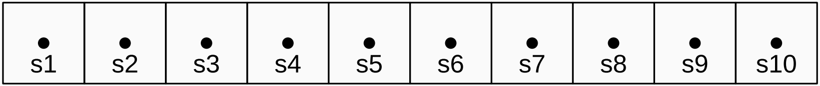
vs.

$$y_t | y_{t-1} \sim \mathcal{N}(\delta + \phi\, y_{t-1},\ \sigma^2)$$

In the case of time, both of these definitions result in the same multivariate distribution for $\mathbf{y}$.

AR in Space

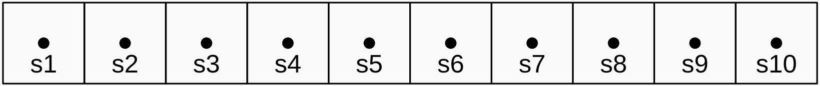




Even in the simplest spatial case there is no clear / unique ordering,

$$
\begin{aligned}
f(y(s_1), \dots, y(s_{10})) &= f(y(s_1))\, f(y(s_2)|y(s_1)) \,\cdots\, f(y(s_{10}|y(s_9), y(s_8), \dots, y(s_1)) \\
&= f(y(s_{10}))\, f(y(s_9)|y(s_{10})) \,\cdots\, f(y(s_1|y(s_2), y(s_3), \dots, y(s_{10})) \\
&= \; ?
\end{aligned}
$$

AR in Space



Even in the simplest spatial case there is no clear / unique ordering,

$$
\begin{aligned}
f(y(s_1), \dots, y(s_{10})) &= f(y(s_1))\, f(y(s_2)|y(s_1)) \cdots f(y(s_{10}|y(s_9), y(s_8), \dots, y(s_1)) \\
&= f(y(s_{10}))\, f(y(s_9)|y(s_{10})) \cdots f(y(s_1|y(s_2), y(s_3), \dots, y(s_{10})) \\
&= \,?
\end{aligned}
$$

Instead we need to think about things in terms of their neighbors / neighborhoods. We define $N(s_i)$ to be the set of neighbors of location s_i.

- If we define the neighborhood based on “touching” then $N(s_3) = \{s_2, s_4\}$
- If we use distance within 2 units then $N(s_3) = \{s_1, s_2, s_3, s_4\}$

Defining the Spatial AR model

Here we will consider a simple average of neighboring observations, just like with the temporal AR model we have two options in terms of defining the autoregressive process,

Defining the Spatial AR model

Here we will consider a simple average of neighboring observations, just like with the temporal AR model we have two options in terms of defining the autoregressive process,

- Simultaneous Autogressve (SAR)

$$y(s) = \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2)$$

Defining the Spatial AR model

Here we will consider a simple average of neighboring observations, just like with the temporal AR model we have two options in terms of defining the autoregressive process,

- Simultaneous Autogressve (SAR)

$$y(s) = \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2)$$

- Conditional Autoregressive (CAR)

$$y(s)|\mathbf{y}(-s) \sim \mathcal{N}\left(\delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s'),\ \sigma^2\right)$$

Simultaneous Autogressve (SAR)

Using

$$y(s) = \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2)$$

we want to find the distribution of $\mathbf{y} = \Big(y(s_1),\, y(s_2),\, \ldots,\, y(s_n)\Big)^t$.

Using

$$y(s) = \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2)$$

we want to find the distribution of $\mathbf{y} = \Big(y(s_1),\, y(s_2),\, \ldots,\, y(s_n)\Big)^t$.

First we can define a weight matrix $\mathbf{W}$ where

$$\{\mathbf{W}\}_{ij} = \begin{cases} 1/|N(s_i)| & \text{if } j \in N(s_i) \\ 0 & \text{otherwise} \end{cases}$$

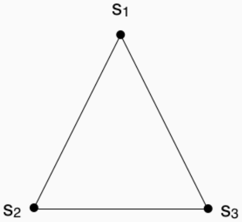
then we can write $\mathbf{y}$ as follows,

$$\mathbf{y} = \phi\, \mathbf{W}\, \mathbf{y} + \boldsymbol{\epsilon}$$

where

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\, \mathbf{I})$$

A toy example



Back to SAR

$$\mathbf{y} = \phi \, \mathbf{W} \, \mathbf{y} + \boldsymbol{\epsilon}$$

Conditional Autogressve (CAR)

This is a bit trickier, in the case of the temporal AR process we actually went from joint distribution $\rightarrow$ conditional distributions (which we were then able to simplify).

Since we don't have a natural ordering we can't get away with this (at least not easily).

Going the other way, conditional distributions $\rightarrow$ joint distribution is difficult because it is possible to specify conditional distributions that lead to an improper joint distribution.

Brooks' Lemma

For sets of observations $\mathbf{x}$ and $\mathbf{y}$ where $p(x) > 0 \;\; \forall\, x \in \mathbf{x}$ and $p(y) > 0 \;\; \forall\, y \in \mathbf{y}$ then

$$\begin{aligned}\frac{p(\mathbf{y})}{p(\mathbf{x})} &= \prod_{i=1}^{n} \frac{p(y_i \mid y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)}{p(x_i \mid y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)} \\ &= \prod_{i=1}^{n} \frac{p(y_i \mid x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_n)}{p(x_i \mid x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_n)}\end{aligned}$$

A simplified example

Let $\mathbf{y} = (y_1, y_2)$ and $\mathbf{x} = (x_1, x_2)$ then we can derive Brook's Lemma for this case,

$$p(y_1, y_2) = p(y_1|y_2)p(y_2)$$

A simplified example

Let $\mathbf{y} = (y_1, y_2)$ and $\mathbf{x} = (x_1, x_2)$ then we can derive Brook's Lemma for this case,

$$
\begin{aligned}
p(y_1, y_2) &= p(y_1|y_2)p(y_2) \\
&= p(y_1|y_2)\frac{p(y_2|x_1)}{p(x_1|y_2)}p(x_1)
\end{aligned}
$$

A simplified example

Let $\mathbf{y} = (y_1, y_2)$ and $\mathbf{x} = (x_1, x_2)$ then we can derive Brook's Lemma for this case,

$$\begin{aligned} p(y_1, y_2) &= p(y_1|y_2)p(y_2) \\ &= p(y_1|y_2)\frac{p(y_2|x_1)}{p(x_1|y_2)}p(x_1) = \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\, p(x_1) \end{aligned}$$

A simplified example

Let $\mathbf{y} = (y_1, y_2)$ and $\mathbf{x} = (x_1, x_2)$ then we can derive Brook's Lemma for this case,

$$\begin{aligned}
p(y_1, y_2) &= p(y_1|y_2)p(y_2) \\
&= p(y_1|y_2)\frac{p(y_2|x_1)}{p(x_1|y_2)}p(x_1) = \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\,p(x_1) \\
&= \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\,p(x_1)\left(\frac{p(x_2|x_1)}{p(x_2|x_1)}\right)
\end{aligned}$$

A simplified example

Let $\mathbf{y} = (y_1, y_2)$ and $\mathbf{x} = (x_1, x_2)$ then we can derive Brook's Lemma for this case,

$$
\begin{aligned}
p(y_1, y_2) &= p(y_1|y_2)p(y_2) \\
&= p(y_1|y_2)\frac{p(y_2|x_1)}{p(x_1|y_2)}p(x_1) = \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\,p(x_1) \\
&= \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\,p(x_1)\left(\frac{p(x_2|x_1)}{p(x_2|x_1)}\right) \\
&= \frac{p(y_1|y_2)}{p(x_1|y_2)}\frac{p(y_2|x_1)}{p(x_2|x_1)}\,p(x_1, x_2)
\end{aligned}
$$

A simplified example

Let $\mathbf{y} = (y_1, y_2)$ and $\mathbf{x} = (x_1, x_2)$ then we can derive Brook's Lemma for this case,

$$
\begin{aligned}
p(y_1, y_2) &= p(y_1|y_2)p(y_2) \\
&= p(y_1|y_2)\frac{p(y_2|x_1)}{p(x_1|y_2)}p(x_1) = \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\,p(x_1) \\
&= \frac{p(y_1|y_2)}{p(x_1|y_2)}p(y_2|x_1)\,p(x_1)\left(\frac{p(x_2|x_1)}{p(x_2|x_1)}\right) \\
&= \frac{p(y_1|y_2)}{p(x_1|y_2)}\,\frac{p(y_2|x_1)}{p(x_2|x_1)}\,p(x_1, x_2)
\end{aligned}
$$

$$
\frac{p(y_1, y_2)}{p(x_1, x_2)} = \frac{p(y_1|y_2)}{p(x_1|y_2)}\,\frac{p(y_2|x_1)}{p(x_2|x_1)}
$$

Utility?

Lets repeat that last example but consider the case where $\mathbf{y} = (y_1, y_2)$ but now we let $\mathbf{x} = (y_1 = 0, y_2 = 0)$

$$\frac{p(y_1, y_2)}{p(x_1, x_2)} = \frac{p(y_1, y_2)}{p(y_1 = 0, y_2 = 0)}$$

Utility?

Lets repeat that last example but consider the case where $\mathbf{y} = (y_1, y_2)$ but now we let $\mathbf{x} = (y_1 = 0, y_2 = 0)$

$$\frac{p(y_1, y_2)}{p(x_1, x_2)} = \frac{p(y_1, y_2)}{p(y_1 = 0, y_2 = 0)}$$

$$p(y_1, y_2) = \frac{p(y_1|y_2)}{p(y_1 = 0|y_2)} \frac{p(y_2|y_1 = 0)}{p(y_2 = 0|y_1 = 0)} \, p(y_1 = 0, y_2 = 0)$$

Utility?

Lets repeat that last example but consider the case where $\mathbf{y} = (y_1, y_2)$ but now we let $\mathbf{x} = (y_1 = 0, y_2 = 0)$

$$\frac{p(y_1, y_2)}{p(x_1, x_2)} = \frac{p(y_1, y_2)}{p(y_1 = 0, y_2 = 0)}$$

$$p(y_1, y_2) = \frac{p(y_1|y_2)}{p(y_1 = 0|y_2)} \frac{p(y_2|y_1 = 0)}{p(y_2 = 0|y_1 = 0)} p(y_1 = 0, y_2 = 0)$$

$$\begin{aligned} p(y_1, y_2) &\propto \frac{p(y_1|y_2)\, p(y_2|y_1 = 0)}{p(y_1 = 0|y_2)} \\ &\propto \frac{p(y_2|y_1)\, p(y_1|y_2 = 0)}{p(y_2 = 0|y_1)} \end{aligned}$$

As applied to a simple CAR

$$y(s_1)|y(s_2) \sim \mathcal{N}(\phi W_{12}\, y(s_2), \sigma^2)$$
$$y(s_2)|y(s_1) \sim \mathcal{N}(\phi W_{21}\, y(s_1), \sigma^2)$$

As applied to a simple CAR

$$y(s_1)|y(s_2) \sim \mathcal{N}(\phi W_{12}\, y(s_2), \sigma^2)$$
$$y(s_2)|y(s_1) \sim \mathcal{N}(\phi W_{21}\, y(s_1), \sigma^2)$$

$$p(y(s_1), y(s_2)) \propto \frac{p(y(s_1)|y(s_2))\; p(y(s_2)|y(s_1) = 0)}{p(y(s_1) = 0|y(s_2))}$$

As applied to a simple CAR

$$y(s_1)|y(s_2) \sim \mathcal{N}(\phi W_{12}\, y(s_2), \sigma^2)$$
$$y(s_2)|y(s_1) \sim \mathcal{N}(\phi W_{21}\, y(s_1), \sigma^2)$$

$$\begin{aligned} p(y(s_1), y(s_2)) &\propto \frac{p(y(s_1)|y(s_2))\; p(y(s_2)|y(s_1) = 0)}{p(y(s_1) = 0|y(s_2))} \\ &\propto \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y(s_1) - \phi\, W_{12}\, y(s_2)\right)^2\right) \exp\left(-\frac{1}{2\sigma^2}\left(y(s_2) - \phi\, W_{21}\, 0\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0 - \phi W_{12} y(s_2)\right)^2\right)} \end{aligned}$$

As applied to a simple CAR

$$y(s_1)|y(s_2) \sim \mathcal{N}(\phi W_{12}\, y(s_2), \sigma^2)$$
$$y(s_2)|y(s_1) \sim \mathcal{N}(\phi W_{21}\, y(s_1), \sigma^2)$$

$$
\begin{aligned}
p(y(s_1), y(s_2)) &\propto \frac{p(y(s_1)|y(s_2))\, p(y(s_2)|y(s_1)=0)}{p(y(s_1)=0|y(s_2))} \\
&\propto \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y(s_1) - \phi\, W_{12}\, y(s_2)\right)^2\right)\exp\left(-\frac{1}{2\sigma^2}\left(y(s_2) - \phi\, W_{21}\, 0\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0 - \phi W_{12} y(s_2)\right)^2\right)} \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left(\left(y(s_1) - \phi\, W_{12}\, y(s_2)\right)^2 + y(s_2)^2 - (\phi W_{21} y(s_2))^2\right)\right)
\end{aligned}
$$

As applied to a simple CAR

$$y(s_1)|y(s_2) \sim \mathcal{N}(\phi W_{12}\, y(s_2), \sigma^2)$$
$$y(s_2)|y(s_1) \sim \mathcal{N}(\phi W_{21}\, y(s_1), \sigma^2)$$

$$\begin{aligned}
p(y(s_1), y(s_2)) &\propto \frac{p(y(s_1)|y(s_2))\, p(y(s_2)|y(s_1)=0)}{p(y(s_1)=0|y(s_2))} \\
&\propto \frac{\exp\left(-\frac{1}{2\sigma^2}\,(y(s_1) - \phi\, W_{12}\, y(s_2))^2\right) \exp\left(-\frac{1}{2\sigma^2}\,(y(s_2) - \phi\, W_{21}\, 0)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\,(0 - \phi W_{12} y(s_2))^2\right)} \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left((y(s_1) - \phi\, W_{12}\, y(s_2))^2 + y(s_2)^2 - (\phi W_{21} y(s_2))^2\right)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left(y(s_1)^2 - \phi\, W_{12}\, y(s_1)\, y(s_2) - \phi\, W_{21}\, y(s_1)\, y(s_2) + y(s_2)^2\right)\right)
\end{aligned}$$

As applied to a simple CAR

$$y(s_1)|y(s_2) \sim \mathcal{N}(\phi W_{12}\, y(s_2), \sigma^2)$$
$$y(s_2)|y(s_1) \sim \mathcal{N}(\phi W_{21}\, y(s_1), \sigma^2)$$

$$\begin{aligned}
p(y(s_1), y(s_2)) &\propto \frac{p(y(s_1)|y(s_2))\; p(y(s_2)|y(s_1)=0)}{p(y(s_1)=0|y(s_2))} \\
&\propto \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y(s_1)-\phi\, W_{12}\, y(s_2)\right)^2\right)\exp\left(-\frac{1}{2\sigma^2}\left(y(s_2)-\phi\, W_{21}\, 0\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0-\phi W_{12} y(s_2)\right)^2\right)} \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left(\left(y(s_1)-\phi\, W_{12}\, y(s_2)\right)^2 + y(s_2)^2 - (\phi W_{21} y(s_2))^2\right)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left(y(s_1)^2 - \phi\, W_{12}\, y(s_1)\, y(s_2) - \phi\, W_{21}\, y(s_1)\, y(s_2) + y(s_2)^2\right)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-0)\begin{pmatrix} 1 & -\phi W_{12} \\ -\phi W_{21} & 1 \end{pmatrix}(\mathbf{y}-0)^t\right)
\end{aligned}$$

Implications for $\mathbf{y}$

$$\boldsymbol{\mu} = 0$$

Implications for $\mathbf{y}$

$$\boldsymbol{\mu} = 0$$

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\phi W_{12} \\ -\phi W_{21} & 1 \end{pmatrix} \\ &= \frac{1}{\sigma^2}(\mathbf{I} - \phi\,\mathbf{W})\end{aligned}$$

Implications for $\mathbf{y}$

$$\boldsymbol{\mu} = 0$$

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\phi W_{12} \\ -\phi W_{21} & 1 \end{pmatrix} \\ &= \frac{1}{\sigma^2}(\mathbf{I} - \phi\,\mathbf{W})\end{aligned}$$

$$\boldsymbol{\Sigma} = \sigma^2(\mathbf{I} - \phi\,\mathbf{W})^{-1}$$

Implications for $\mathbf{y}$

$$\boldsymbol{\mu} = 0$$

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\phi W_{12} \\ -\phi W_{21} & 1 \end{pmatrix} \\ &= \frac{1}{\sigma^2}(\mathbf{I} - \phi\,\mathbf{W})\end{aligned}$$

$$\boldsymbol{\Sigma} = \sigma^2(\mathbf{I} - \phi\,\mathbf{W})^{-1}$$

we can then conclude that for $\mathbf{y} = (y(s_1),\ y(s_2))^t$,

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{0},\ \sigma^2(\mathbf{I} - \phi\,\mathbf{W})^{-1}\right)$$

which generalizes for all mean 0 CAR models.

General Proof

Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook's lemma,

$$\frac{p(\mathbf{y})}{p(\mathbf{0})} = \prod_{i=1}^{n} \frac{p(y_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}$$

Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook's lemma,

$$\begin{aligned}\frac{p(\mathbf{y})}{p(\mathbf{0})} &= \prod_{i=1}^{n} \frac{p(y_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)} \\ &= \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}\end{aligned}$$

Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook's lemma,

$$\frac{p(\mathbf{y})}{p(\mathbf{0})} = \prod_{i=1}^{n} \frac{p(y_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}$$

$$= \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \phi \sum_{j<i} W_{ij}\, y_j\right)^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\phi \sum_{j<i} W_{ij}\, y_j\right)^2\right)$$

General Proof

Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook's lemma,

$$
\begin{aligned}
\frac{p(\mathbf{y})}{p(\mathbf{0})} &= \prod_{i=1}^{n} \frac{p(y_i|y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i|y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)} \\
&= \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)} \\
&= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \phi \sum_{j<i} W_{ij}\, y_j\right)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\phi \sum_{j<i} W_{ij}\, y_j\right)^2\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 - 2\phi y_i \sum_{j<i} W_{ij}\, y_j\right)
\end{aligned}
$$

Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook's lemma,

$$
\begin{aligned}
\frac{p(\mathbf{y})}{p(\mathbf{0})} &= \prod_{i=1}^{n} \frac{p(y_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)} \\
&= \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)} \\
&= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \phi \sum_{j<i} W_{ij}\, y_j\right)^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\phi \sum_{j<i} W_{ij}\, y_j\right)^2\right) \\
&= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 - 2\phi y_i \sum_{j<i} W_{ij}\, y_j\right) \\
&= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 - \phi \sum_{i=1}^{n} \sum_{j=1}^{n} y_i\, W_{ij}\, y_j\right) \quad (\text{if } W_{ij} = W_{ji})
\end{aligned}
$$

Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook's lemma,

$$
\begin{aligned}
\frac{p(\mathbf{y})}{p(\mathbf{0})} &= \prod_{i=1}^{n} \frac{p(y_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i | y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)} \\
&= \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2\sigma^2}\left(y_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\left(0_i - \phi \sum_{j<i} W_{ij}\, y_j - \phi \sum_{j>i} 0_j\right)^2\right)} \\
&= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \phi \sum_{j<i} W_{ij}\, y_j\right)^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\phi \sum_{j<i} W_{ij}\, y_j\right)^2\right) \\
&= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 - 2\phi y_i \sum_{j<i} W_{ij}\, y_j\right) \\
&= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 - \phi \sum_{i=1}^{n} \sum_{j=1}^{n} y_i\, W_{ij}\, y_j\right) \quad (\text{if } W_{ij} = W_{ji}) \\
&= \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - 0)^t (\mathbf{I} - \phi \mathbf{W})(\mathbf{y} - 0)\right)
\end{aligned}
$$

- Simultaneous Autogressve (SAR)

$$y(s) = \phi \sum_{s'} W_{s\,s'}\, y(s') + \epsilon$$

$$\mathbf{y} \sim \mathcal{N}(0,\ \sigma^2\, ((\mathbf{I} - \phi \mathbf{W})^{-1})((\mathbf{I} - \phi \mathbf{W})^{-1})^t)$$

- Conditional Autoregressive (CAR)

$$y(s)|\mathbf{y}(-s) \sim \mathcal{N}\left(\sum_{s'} W_{s\,s'}\, y(s'),\ \sigma^2\right)$$

$$\mathbf{y} \sim \mathcal{N}(0,\ \sigma^2\, (\mathbf{I} - \phi \mathbf{W})^{-1})$$

Generalization

- Adopting different weight matrices, $\mathbf{W}$
  - Between SAR and CAR model we move to a generic weight matrix definition (beyond average of nearest neighbors)
  - In time we varied p in the $AR(p)$ model, in space we adjust the weight matrix.
  - In general having a symmetric W is helpful, but not required

Generalization

- Adopting different weight matrices, $\mathbf{W}$
  - Between SAR and CAR model we move to a generic weight matrix definition (beyond average of nearest neighbors)
  - In time we varied p in the $AR(p)$ model, in space we adjust the weight matrix.
  - In general having a symmetric W is helpful, but not required
- More complex Variance (beyond $\sigma^2\, I$)
  - σ^2 can be a vector (differences between areal locations)
  - i.e. since areal data tends to be aggregated - adjust variance based on sample size
  - i.e. scale based on the number of neighbors

Some specific generalizations

Generally speaking we will want to work with a scaled / normalized version of the weight matrix,

$$\frac{W_{ij}}{W_{i\cdot}}$$

When W is derived from an adjacency matrix, $\mathbf{A}$, we can express this as

$$\mathbf{W} = \mathbf{D}^{-1}\mathbf{A}$$

where $\mathbf{D}^{-1} = \text{diag}(1/|N(s_i)|)$.

We can also allow σ^2 to vary between locations, we can define this as $\mathbf{D}_{\sigma^2} = \text{diag}(\sigma_i^2)$ and most often we use

$$\mathbf{D}_{\sigma^2}^{-1} = \text{diag}\left(\frac{\sigma^2}{|N(s_i)|}\right) = \sigma^2\mathbf{D}^{-1}.$$

Revised SAR Model

- Formula Model

$$y(s_i) = X_{i\cdot}\beta + \phi \sum_{j=1}^{n} D_{jj}^{-1} \, A_{ij} \left(y(s_j) - X_{j\cdot}\beta\right) + \epsilon_i$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\, \mathbf{D}_{\sigma^2}^{-1}) = \mathcal{N}(\mathbf{0},\, \sigma^2\mathbf{D}^{-1})$$

- Joint Model

Revised SAR Model

- Formula Model

$$y(s_i) = X_{i\cdot}\beta + \phi \sum_{j=1}^{n} D_{jj}^{-1} \, A_{ij} \left(y(s_j) - X_{j\cdot}\beta\right) + \epsilon_i$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\, \mathbf{D}_{\sigma^2}^{-1}) = \mathcal{N}(\mathbf{0},\, \sigma^2\mathbf{D}^{-1})$$

- Joint Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \phi\mathbf{D}^{-1}\mathbf{A}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) + \boldsymbol{\epsilon}$$

Revised SAR Model

- Formula Model

$$y(s_i) = X_{i\cdot}\beta + \phi \sum_{j=1}^{n} D_{jj}^{-1} A_{ij} \left(y(s_j) - X_{j\cdot}\beta\right) + \epsilon_i$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\, \mathbf{D}_{\sigma^2}^{-1}) = \mathcal{N}(\mathbf{0},\, \sigma^2 \mathbf{D}^{-1})$$

- Joint Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \phi \mathbf{D}^{-1}\mathbf{A}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) + \boldsymbol{\epsilon}$$

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, (\mathbf{I} - \phi\mathbf{D}^{-1}\mathbf{A})^{-1}\sigma^2\mathbf{D}^{-1}\left((\mathbf{I} - \phi\mathbf{D}^{-1}\mathbf{W})^{-1}\right)^{t}\right)$$

Revised CAR Model

- Conditional Model

$$y(s_i)|\mathbf{y}_{-s_i} \sim \mathcal{N}\left(X_{i\cdot}\beta + \phi \sum_{j=1}^{n} \frac{W_{ij}}{D_{ii}}\,(y(s_j) - X_{j\cdot}\beta),\; \sigma^2 D_{ii}^{-1}\right)$$

- Joint Model

Revised CAR Model

- Conditional Model

$$y(s_i)|\mathbf{y}_{-s_i} \sim \mathcal{N}\left(X_{i\cdot}\beta + \phi \sum_{j=1}^{n} \frac{W_{ij}}{D_{ii}} \left(y(s_j) - X_{j\cdot}\beta\right),\ \sigma^2 D_{ii}^{-1}\right)$$

- Joint Model

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta},\ \Sigma_{CAR})$$

Revised CAR Model

- Conditional Model

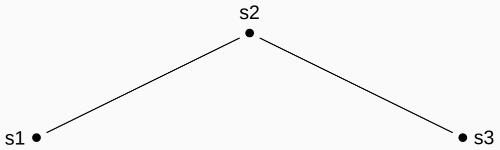
$$y(s_i)|\mathbf{y}_{-s_i} \sim \mathcal{N}\left(X_{i\cdot}\beta + \phi\sum_{j=1}^{n}\frac{W_{ij}}{D_{ii}}\,(y(s_j) - X_{j\cdot}\beta),\; \sigma^2 D_{ii}^{-1}\right)$$

- Joint Model

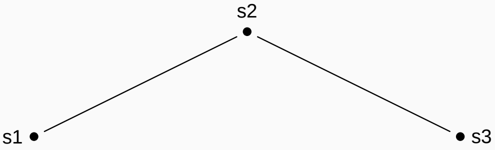
$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta},\; \Sigma_{CAR})$$

$$\begin{aligned}\Sigma_{CAR} &= (\mathbf{D}_\sigma\,(\mathbf{I} - \phi\mathbf{D}^{-1}\mathbf{W}))^{-1}\\ &= (1/\sigma^2\mathbf{D}\,(\mathbf{I} - \phi\mathbf{D}^{-1}\mathbf{W}))^{-1}\\ &= (1/\sigma^2(\mathbf{D} - \phi\mathbf{W}))^{-1}\\ &= \sigma^2(\mathbf{D} - \phi\mathbf{W})^{-1}\end{aligned}$$

Toy CAR Example



Toy CAR Example



$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad \mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

. . .

$$\mathbf{\Sigma} = \sigma^2 \left(\mathbf{D} - \phi\,\mathbf{W}\right) = \sigma^2 \begin{pmatrix} 1 & -\phi & 0 \\ -\phi & 2 & -\phi \\ 0 & -\phi & 1 \end{pmatrix}^{-1}$$

When does Σ exist?

```
check_sigma = function(phi) {
  Sigma_inv = matrix(c(1,-phi,0,-phi,2,-phi,0,-phi,1), ncol=3, byrow=TRUE)
  solve(Sigma_inv)
}

check_sigma(phi=0)
##      [,1] [,2] [,3]
## [1,]    1  0.0    0
## [2,]    0  0.5    0
## [3,]    0  0.0    1

check_sigma(phi=0.5)
##           [,1]      [,2]      [,3]
## [1,] 1.1666667 0.3333333 0.1666667
## [2,] 0.3333333 0.6666667 0.3333333
## [3,] 0.1666667 0.3333333 1.1666667

check_sigma(phi=-0.6)
##          [,1]     [,2]     [,3]
## [1,]  1.28125 -0.46875  0.28125
## [2,] -0.46875  0.78125 -0.46875
## [3,]  0.28125 -0.46875  1.28125
```

```r
check_sigma(phi=1)
## Error in solve.default(Sigma_inv): Lapack routine dgesv: system is exactl

check_sigma(phi=-1)
## Error in solve.default(Sigma_inv): Lapack routine dgesv: system is exactl

check_sigma(phi=1.2)
##            [,1]      [,2]       [,3]
## [1,] -0.6363636 -1.363636 -1.6363636
## [2,] -1.3636364 -1.136364 -1.3636364
## [3,] -1.6363636 -1.363636 -0.6363636

check_sigma(phi=-1.2)
##            [,1]      [,2]       [,3]
## [1,] -0.6363636  1.363636 -1.6363636
## [2,]  1.3636364 -1.136364  1.3636364
## [3,] -1.6363636  1.363636 -0.6363636
```

When is Σ positive semidefinite?

```
check_sigma_pd = function(phi) {
  Sigma_inv = matrix(c(1,-phi,0,-phi,2,-phi,0,-phi,1), ncol=3, byrow=TRUE)
  chol(solve(Sigma_inv))
}

check_sigma_pd(phi=0)
##      [,1]      [,2] [,3]
## [1,]    1 0.0000000    0
## [2,]    0 0.7071068    0
## [3,]    0 0.0000000    1

check_sigma_pd(phi=0.5)
##          [,1]      [,2]      [,3]
## [1,] 1.080123 0.3086067 0.1543033
## [2,] 0.000000 0.7559289 0.3779645
## [3,] 0.000000 0.0000000 1.0000000

check_sigma_pd(phi=-0.6)
##          [,1]       [,2]       [,3]
## [1,] 1.131923 -0.4141182  0.2484709
## [2,] 0.000000  0.7808688 -0.4685213
## [3,] 0.000000  0.0000000  1.0000000
```

```r
check_sigma_pd(phi=1)
## Error in solve.default(Sigma_inv): Lapack routine dgesv: system is exactl

check_sigma_pd(phi=-1)
## Error in solve.default(Sigma_inv): Lapack routine dgesv: system is exactl

check_sigma_pd(phi=1.2)
## Error in chol.default(solve(Sigma_inv)): the leading minor of order 1 is

check_sigma_pd(phi=-1.2)
## Error in chol.default(solve(Sigma_inv)): the leading minor of order 1 is
```

Generally speaking just like the AR(1) model for time series we require that $|\phi| < 1$ for the CAR model to be proper.

These results for ϕ also apply in the context where σ_i^2 is constant across locations (i.e. $\mathbf{\Sigma} = \left(\sigma^2 \left(\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{W}\right)\right)^{-1}$)

As a side note, the special case where $\phi = 1$ is known as an intrinsic autoregressive (IAR) model and they are popular as an *improper* prior for spatial random effects. An additional sum constraint is necessary for identifiability ($\sum_{i=1}^{n} y(s_i) = 0$).